Published as: *Cell*. 2011 June 10; 145(6): 981–992.

Metazoan operons accelerate recovery from growth arrested states

Alon Zaslaver[1], **L. Ryan Baugh**[2], and **Paul W. Sternberg**[1]

[1]Howard Hughes Medical Institute and Division of Biology, California Institute of Technology, 1200 E. California Blvd., Pasadena, California 91125

[2]Department of Biology and IGSP Center for Systems Biology, Duke University, Box 90338, Durham, NC 27708

Summary

Existing theories explain why operons are advantageous in prokaryotes, but their occurrence in metazoans is an enigma. Nematode operon genes, typically consisting of growth genes, are significantly up-regulated during recovery from growth-arrested states. This expression pattern is anti-correlated to non-operon genes consistent with a competition for transcriptional resources. We find that transcriptional resources are initially limiting during recovery, and that recovering animals are highly sensitive to any additional decrease in transcriptional resources. Operons become advantageous because by clustering growth genes into operons, fewer promoters compete for the limited transcriptional machinery, effectively increasing the concentration of transcriptional resources, and accelerating recovery. Mathematical modeling reveals how a moderate increase in transcriptional resources can substantially enhance transcription rate and recovery. This design principle occurs in different nematodes and the chordate *C. intestinalis*. As transition from arrest to rapid growth is shared by many metazoans, operons could have evolved to facilitate these processes.

Introduction

Operons constitute a unique gene organization where two or more genes are transcribed together on the same mRNA unit (Jacob et al., 1960). Operons are found ubiquitously in prokaryotic genomes and are thought to be advantageous because they provide co-regulation of related genes (Price et al., 2005), and allow successful horizontal gene transfers of related genes often clustered in the same operon (‘selfish operons’ (Lawrence, 2003)). As new species are added to the growing list of sequenced genomes, it is has become apparent that various metazoans contain operons as well. For example, in the nematode phylum ~20% of the transcriptome is found in operons (Abad et al., 2008; Blumenthal et al., 2002; Blumenthal and Gleason, 2003; Ghedin et al., 2007; Guiliano and Blaxter, 2006; Opperman et al., 2008; Qian and Zhang, 2008; Stein et al., 2003; Zorio et al., 1994), and similar fraction of operons is also found in the chordates *C. intestinalis* (Satou et al., 2008) and *O.*

Corresponding author - Paul Sternberg (pws@caltech.edu).

Author contributions: AZ conceived the idea of this study, designed and performed all of the experiments, the mathematical modeling, and carried out all the data and bioinformatics analyses. LRB contributed experimental tools, shared unpublished data and helped analyze and interpret the data. AZ and PWS wrote the paper.

dioica (Ganot et al., 2004). Could similar evolutionary forces drive the propagation of operons in metazoans? As horizontal gene transfer is not known to occur between metazoans, operons are thought to evolve to allow co-regulation. Indeed, co-regulation among operon genes in *C. elegans* was observed; however, the co-regulation was found to be weak and only slightly higher than the co-regulation of two neighboring genes that do not form an operon (Chen and Stein, 2006; Lercher et al., 2003). Moreover, most operons contain genes that belong to different functional pathways, so the need for their co-regulation remains elusive.

A recent study showed that operon genes are highly enriched in the set of genes expressed in the *C. elegans* germline tissue (Reinke and Cutter, 2009). This suggests that operon genes play a critical role in germline function, but what that role is still unknown.

Here we address the question of why operons would evolve in metazoans, and present a novel unifying theory supported by multiple independent findings to explain this puzzling genomic organization. Our findings demonstrate that operon genes, comprising highly expressed growth genes, are required for rapid recovery from growth arrested states into a fast growing state. We also show that transcriptional resources are limited during recovery from growth arrested states. Operons thus become advantageous because clustering abundant genes into operons increases the concentration of the limited transcriptional resources as fewer promoters compete for these limited resources. All together, we provide evidence that operons could evolve as an evolutionary solution to increase recovery rate from growth arrested states. We find this evolutionary solution to arise in various metazoans that undergo growth arrest periods as part of their lifecycle.

Results

Operons comprise highly expressed growth genes

We begin by asking whether the set of thousands of operon genes in *C. elegans* (Blumenthal et al., 2002) share a common functional relationship. Analyzing which biological processes are enriched with operon genes (GO annotations, WormBase WS190), we found that growth-related processes are highly enriched with operon genes (Table S1). This observation is consistent with previous findings that many ‘housekeeping’ genes are found in operons (Blumenthal and Gleason, 2003). These genes function in a wide-range of growth related processes during the worms’ lifecycle from optimal larval growth to germline production (Fig 1A). Since operon genes are typically growth related genes (88% of the operon genes are growth related, Table S1), we analyzed their expression pattern throughout the worms’ lifecycle using existing microarray data (http://elegans.bcgsc.bc.ca/). We found that the average expression level of operon genes is about two-fold higher compared to the average expression level of non-operon genes (Fig 1A). These observations are also consistent with the protein levels of these genes (Schrimpf et al., 2009). The higher expression levels are consistent throughout all stages of growth: embryos through all larval stages to adults.

The expression of operon genes is up regulated upon recovery from growth arrested states, and is anti-correlated to the expression of non-operon genes

When worms encounter unfavorable conditions they do not follow the canonical developmental path but instead enter into particular growth arrested states: notably L1 larval stage arrest and dauer arrest (Johnson et al., 1984; Klass and Hirsh, 1976) (Fig 1A). Dauer arrest is particularly ecologically relevant as it is the predominant state in which worms are found in nature (Barriere and Felix, 2005). Once conditions improve, the worms recover, resume growth, and complete larval development. Since operons contain highly-expressed

growth genes, one would expect them to play a fundamental role in allowing fast and efficient recovery into a growing state, as predicted by (Blumenthal and Gleason, 2003).

We therefore analyzed expression dynamics of operon genes upon recovery from both L1 and dauer arrest using published microarray data (Wang and Kim, 2003). We find a significant difference between the expression patterns of operon and non-operon genes (Fig 1B-C). During the recovery from both L1 arrest and dauer arrest, operon genes start at low levels and rapidly increase, while non-operon genes show relatively constant expression levels with a mild decrease. Moreover, we find a significant anti-correlation in the expression pattern of the two sets of genes: correlation coefficients of −0.59 and −0.78 for recovery from L1 arrest and dauer, respectively. The anti-correlation becomes more pronounced if non-operon growth genes are excluded (Fig S1). Note that the first step in recovery from dauer is characterized by a transient period in which growth is not yet initiated. During this period, cell adhesion, transport and signaling genes are up regulated probably to prime the recovery process (Cassada and Russell, 1975; Golden and Riddle, 1984; Wang and Kim, 2003). Following this transient period operon genes are rapidly up-regulated to support the concomitant growth processes (Fig 1B).

L1 larvae right after hatching are also considered to undergo a transition similar to recovery from growth arrest, as they initiate feeding and rapid growth having depleted all food resources inside the egg (Fig 1A). We therefore analyzed the expression profile of operon genes following hatching using time-series microarray data (Baugh et al., 2009). Similarly, we find significantly distinct expression profiles for the two sets of genes: expression of operon genes increases while expression of non-operon genes decreases (Fig 1D). As in recovery from L1 and dauer arrest, the two expression profiles are almost completely anti-correlated (correlation coefficient = −0.81).

As fast-growing and mass-expanding stages primarily involve the expression of operon genes, we asked how operon genes are regulated during embryogenesis, which is not associated with growth but rather cell divisions, patterning and morphogenesis (Fig 1A). We analyzed high-temporal-resolution microarray measurements starting in the 4-cell stage, the onset of embryonic transcription, and during the following three hours as cell fates are specified and the embryo is patterned (Baugh et al., 2003). In sharp contrast to arrest recovery, expression of operon genes decreases while expression of non-operon genes moderately increases (Fig 1E). Strikingly, while these expression profiles are the reverse of those seen in growth recovery processes, the anti-correlation in expression profiles is preserved (correlation coefficient = −0.86).

Rapid growth also takes place in the germline, facilitating rapid proliferation of germ cells (Fig 1A). Furthermore, the rate at which hermaphroditic nematodes (e.g., *C. elegans*) reproduce is determined by the rate of oocyte production, and oocytes must be energy-rich to support embryogenesis as embryos do not feed but depend on maternal provisions. Indeed, the vast majority of operon genes are expressed in the *C. elegans* germline (Reinke and Cutter, 2009), presumably to support fast proliferation as well as endowing oocytes with growth gene products to support embryogenesis. The importance of oocyte endowment of growth gene products for early embryogenesis is underscored by the fact that operon genes are down regulated during the first hours of embryogenesis (Fig 1E).

The anti-correlation between operon and non-operon genes, consistently observed in all data sets, is remarkable given that over 2,000 genes make up the averaged expression profile of operon genes and nearly 17,000 genes are averaged to provide the expression pattern of non-operon genes (Fig 1, Fig S1). This anti-correlation suggests that operon and non-operon genes have a zero-sum relationship, as if competing for limited transcriptional resources (Fig

S1E). Furthermore, the vast majority of operon genes are tightly correlated with the mean expression profile of operon genes. In contrast, the vast majority of non-operon genes are anti-correlated with the mean expression profile of operon genes (Fig S3). Importantly, in all data sets, operon genes exhibit a unique and significantly distinctive expression profile (Fig 1, Fig S1) that cannot be recapitulated at random (see bootstrap analysis). This unique expression profile of operon genes is biased neither by ribosomal protein genes, which tend to be clustered into operons (Fig S2A-D), nor by the fact that operons contain highly expressed genes (Fig S2E-F), but is rather a genuine outcome of growth genes being organized into operons.

The expression of operon genes are highly correlated at the beginning of recovery but this correlation decreases over time (Figs S5A-C). In contrast, neighboring genes tend to show a constant correlation score only slightly lower than that of the operon genes, in agreement with previous findings (Chen and Stein, 2006; Lercher et al., 2003). However, random pairs of operon genes (i.e., genes not in the same operon) are highly correlated which may explain why genes on a given operon typically do not share an obvious specific functional relationship, but nevertheless belong to the broader class of growth-related genes, which are collectively co-expressed immediately upon recovery. Indeed, when operon genes are not up-regulated on average, as observed during early embryogenesis (Fig 1E), expressions of random pairs of operon genes are not correlated (Fig S5D).

All these observations demonstrate that operon genes, comprising primarily growth genes, are rapidly up-regulated upon animals' recovery from arrested states into a fast-growing state. As not all growth genes are clustered into operons (e.g., only ~50% of transcriptional machinery is in operons), we also analyzed the expression profiles of genes coding for gene expression machinery (transcription and translation) as well as energy generating genes. We found that, upon recovery, this set of genes highly correlates with the expression profile of operon genes upon recovery (correlation coefficient > 0.9; Fig S4).

Transcriptional resources are limited during the transition from arrest to growth

To allow the massive transcription required during arrest recovery, the transcriptional machinery needs to be fully engaged. However, several lines of evidence indicate that transcriptional resources are limiting during developmental arrest: The most direct evidence is based on run-on transcription assays of nuclei isolated from dauers (Dalley and Golomb, 1992). In these experiments, RNA Pol II transcription levels in dauers were only ~15% of those in non-arrested worms. Furthermore, following one hour of recovery, the transcription rate was increased to only ~23% of that seen in non-arrested worms. Another line of evidence is based on gene expression analysis following recovery from both L1 arrest and dauer arrest (Wang and Kim, 2003) (Fig 2). We find that the majority of the genes encoding transcriptional machinery components are expressed at lower levels during growth arrest (when compared to non-arrested growth), and most of the transcriptional machinery genes are significantly up-regulated upon recovery [i.e., 63% of the genes are down regulated during L1 arrest (66% in dauer) and expression of 88% significantly increases upon recovery from L1 arrest (58% from dauer); Fig 2A,C and data analysis in SI]. Furthermore, the expression of these transcriptional machinery genes increases significantly more than the average increase observed for the overall operon genes (2-3 fold versus 20%, Fig 2B,D).

The low levels of transcriptional resources during growth arrest were probably tuned by evolution to meet the minimal requirement for transcription at that stage. This strategy of storing low levels of transcriptional machines at times that they are not needed can save valuable resources that can then be directed towards other processes that might be more crucial for survival in harsh and limited conditions that induced the growth arrest in the first place. The worm therefore faces an interesting optimization problem: how to maintain low

levels of transcriptional resources when they are not needed (during the arrest), but at the same time to ensure that these low levels will support a fast and efficient transition from arrest into growth, a process that necessitates the rapid transcription of thousands of highly expressed genes.

We suggest that clustering genes into operons has evolved as a strategy to solve this issue: by clustering genes into operons, the number of promoters competing to recruit transcriptional machinery decreases so that more transcriptional resources per promoter become available, increasing the effective concentration of the transcriptional machinery (Fig 3). This strategy is particularly effective since highly expressed growth genes that require a significant fraction of transcriptional machinery are specifically clustered in operons. As transcription is a bursty process (Chubb et al., 2006; Golding et al., 2005), initiation of transcription is a major limiting step. Clustering these highly demanded genes in operons reduces the time required for re-initiating transcription of downstream genes in an operon (See Experimental Procedures for full analysis).

A small increase in the levels of transcriptional machinery can lead to a substantial speed up in transcription rate

For an intuitive and quantitative understanding of the possible transcriptional speed up, we use a simple model based on the Hill function (Alon, 2006), $\beta_{norm}=TM^n/(K^n+TM^n)$ (where β_{norm} is transcription rate normalized to the maximal transcription rate; TM is the concentration of the transcriptional machinery; K denotes the affinity of the machinery to promoter sites, and n is the Hill coefficient representing cooperativity). Assembly of the multi-subunits of the transcriptional complexes is characterized by high cooperativity (n) and is estimated to be greater than one (Alon, 2006; Carey, 1998). The greater the cooperativity, the sharper is the transition from low to high transcription rates (Fig S6A). However, when transcriptional resources are low (i.e., $TM < K$), higher cooperativity yields sharper reduction in transcription rate. It is in this regime where a moderate increase in the levels of transcriptional machinery can significantly accelerate transcription rate (Figs S6B-C). A significant increase of ~100% in transcription rate is predicted for a wide range of possible values of transcriptional machinery levels (TM/K) and cooperativity (n) (Fig 4A, see Experimental Procedures and SI for full analysis).

Recovery from growth arrest involves the production and degradation of many components, and initial levels of transcriptional machinery become a key factor to allow efficient recovery. To understand how a small increase in initial levels of transcriptional resources affects recovery we simulated their accumulation over time and followed the time for transcriptional machinery levels to reach half of their maximal levels (SI; Fig 4B; Fig S6D-F). This simulation predicts that when initial levels of transcriptional machinery are low, recovery is essentially impossible, since degradation of transcriptional components outweighs their production. When initial levels are high, a moderate acceleration in the accumulation of transcriptional machinery is predicted. However, when initial levels are intermediate, the transition from the non-recovery regime to full recovery is predicted to be very sharp. In this intermediate regime a small increase in the initial levels of transcriptional resources can determine whether the organism recovers or dies (Fig 4B, Fig S6F).

Although this is a simplified model (see supplemental experimental procedures, modeling), any increase in the probability or in the rate of recovery is likely to be strongly selected during evolution, and mechanisms that increase the probability of recovery may also extend the amount of time larvae can survive in an arrested state retaining ability to recover.

Animals recovering from growth arrest are highly sensitive to any further decrease in transcriptional resources

We next experimentally studied how limited transcriptional resources affect recovery. As shown in Fig 2, there are many key transcriptional machinery genes that are found at low levels during the arrest and which are substantially up-regulated upon recovery. Thus, these genes are predicted to play a major role in promoting efficient transcription and rapid recovery. As these are essential genes, mutants are generally not viable; however, four such homozygous mutants [*taf-9*(*ok2871Δ*), *taf-11.2(gk682Δ)*, *cit-1.1*(*gk316Δ*) and *cic-1*(*tm3740Δ*)] are viable, possibly because they share functional redundancy with other genes (e.g, *cit-1.1* is cyclin T, which functions redundantly with *cit-1.2*). Strikingly, we found that the growth rates of all four mutants were reduced by 2-3 fold upon recovery from L1 arrest (when compared to control strains and when compared to growth rate in non-arrested conditions, Fig 5A-E). Thus, while lower levels of transcriptional machinery do not significantly affect growth rate in non-arrested conditions, these lower levels significantly reduce growth rate upon recovery.

We also studied how recovery rate is affected when specifically reducing levels of RNA Pol II, using the drug α–amanitin, which binds and inhibits RNA Pol II (Dalley and Golomb, 1992; Sanford et al., 1983). We compared growth rates of arrested L1s and non-arrested L1s treated with the same sub-lethal concentrations of α-amanitin (Fig 6A-C). We find that non-arrested L1s are unaffected by low levels of α-amanitin (1 μg/ml and 2 μg/ml), and grow at similar rate as untreated L1s. However, the same low levels of α–amanitin significantly slow growth of arrested L1s recovering from starvation when compared to recovery in the absence of α–amanitin (Fig 6A-B). At higher levels of α–amanitin (5 μg/ml) growth rate is considerably impaired during recovery from L1 arrest, while only a mild reduction in growth rate observed for non-arrested L1s. Exposing L1 larvae to higher levels of α–amanitin (2 μg/ml and 5 μg/ml) prevents them from completing all larval stages. However, non-arrested L1s advance faster throughout the larval stages, reaching higher stages than arrested L1s, which stop growing at earlier stages. Together, the recovery experiments using transcription mutants and the RNA Pol II inhibitor α-amanitin demonstrate that recovering animals are sensitive to any further limitation of transcriptional components, and, hence, any strategy that increases the effective levels of these components would be strongly selected for during evolution.

Operons evolved in diverse animals in which fast recovery from arrested states is part of their lifecycle

Does the function of operons presented here apply to all nematodes? Free-living nematodes closely related to *C. elegans* show high similarity in operon organization (Hillier et al., 2005; Qian and Zhang, 2008; Stein et al., 2003). For example, *C. elegans* and *C. briggsae* share 96% identity in operon structures (Stein et al., 2003). While similar lifestyle may dictate which *Caenorhabditis* genes would cluster into operons, it is also possible that these species are too close on the evolutionary time scale and did not undergo substantial divergence. We therefore analyzed putative operon genes in two evolutionarily distant nematodes: *Pristionchus pacificus* and *Brugia malayi*. In sharp contrast to the free-living Caenorhabditis nematodes, *P. pacificus* and *B. malayi* occupy different ecological niches: *P. pacificus* is often associated with beetles, where it is found in the dauer stage as long as the beetle is alive. Once the host dies, dauers recover into a growing state by feeding on microbes developed on the host carcass. Interestingly, under lab conditions, dauers survive up to one year, twice as long as *C. elegans* (Mayer and Sommer, 2011). Rapid and efficient recovery is therefore advantageous as dauers compete for transient resources. *B. malayi* is a parasitic nematode passing between mosquitoes (carriers) and humans (hosts) (Ghedin et al., 2007). Its lifecycle consists of two obligatory growth-arrested states analogous to L1 and dauer

arrest in *C. elegans*; *B. malayi* reproduction, thus, depends on recovery from these arrested states, and its fitness presumably depends on the rate of recovery (Fig 7A).

These two species contain putative operons that make up nearly 20% of their genes, very similar to the fraction of operon genes found in *C. elegans* (Blumenthal et al., 2002; Blumenthal and Gleason, 2003; Dieterich et al., 2008; Ghedin et al., 2007) (Fig S7; Experimental Procedures). Despite the ecological differences, operon genes in both *P. pacificus* and *B. malayi* contain primarily growth-related genes (Fig 7B). Strikingly, although only 10% of the operons in *B. malayi* and *C. elegans* are syntenic (Ghedin et al., 2007), over 90% of the operon genes in *B. malayi* with identified orthologs in *C. elegans* are also found in operons in *C. elegans*. Thus, although the two species occupy completely different niches harboring substantial genomic rearrangements, similar genes were evolutionarily selected to be clustered into operons.

Operon structures are also found in other metazoans outside the nematode phylum (e.g. the chordates *C. intestinalis* (Satou et al., 2008) and *O. dioica* (Ganot et al., 2004)). Could the same design principle account for operon formation in a chordate? To address this question, we focused on *C. intestinalis* for which gene expression data throughout its entire lifecycle is available (Azumi et al., 2007). Strikingly, we found that operon genes showed an expression profile, different from the expression profiles of non-operon genes (Fig 7C). Expression of operon genes decreases during embryogenesis and larval stages and increased dramatically right after metamorphosis in the transition from larvae to juvenile. Since metamorphosis of *C. intestinalis* is characterized by the transformation of a non-feeding mobile larva into a filter-feeding growing juvenile, the increase in operon gene expression at metamorphosis could facilitate fast recovery from the non-feeding state. Furthermore, just like in *C. elegans*, the expression profile of operon genes is entirely anti-correlated to the expression profile of non-operon genes (correlation coefficient = −0.96; Fig 7C **inset**). We next analyzed Gene Ontology (GO) terms associated with *C. intestinalis* operon genes. Growth related processes, typically associated with nematode operon genes, are also significantly enriched within *C. intestinalis* operon genes (Fig S7I).

Discussion

Various animals arrest their growth when environmental conditions become unfavorable. Strong selection forces will therefore favor those that can rapidly and efficiently recover back into a growing state. In this study, we present a theory that explains how operons become advantageous to promote fast recovery from growth arrested states. Importantly, our theory of metazoan operon evolution is consistent with previous possible explanations (Blumenthal and Gleason, 2003; Chen and Stein, 2006; Lercher et al., 2003; Qian and Zhang, 2008; Reinke and Cutter, 2009), reconciling them under one unifying model (a detailed discussion of the proposed theories is found in SI).

Recent studies revealed that some operons contain internal promoters (Huang et al., 2007; Whittle et al., 2008). It is possible that individual operon genes could be temporally and spatially regulated independently of operon-wide regulation, possibly modifying their expression. However, when transcriptional resources are limiting and operon genes are required to be up-regulated during arrest recovery, polycistronic expression would be advantageous and presumably predominate over the use of internal promoters, as also manifested by their collective co-regulation (Fig S5).

The model presented here raises several predictions that would be valuable to test. An intriguing experiment to carry out would involve the replacement of several operons by their individual genes where each gene is controlled by its own promoter. While a significant

amount of operons will probably need to be replaced, the prediction is that these mutants will recover slower than the wild type. In addition, it will be interesting see how general is this evolutionary solution. As more metazoan genomes become available, it will be possible to assess whether this genomic organization is universal across all animals that go through growth arrest as part of their life cycle.

The proposed theory may also be relevant to prokaryotes, as they often face growth-limiting conditions that induce entry into growth arrested states (i.e. in the form of spores or entrance into stationary phase (Lewis, 2007)). Similarly, various global organizations of gene expression patterns evolved to support the fast growth of bacteria. In particular, when resources are limited, allocation of translational resources is quantitatively coupled to the growth rate (Scott et al., 2010; Zaslaver et al., 2009). In future, it would be interesting to see if similar strategies evolved in animals that undergo fast recovery from growth arrested states.

Transcriptional machinery is probably not the only component limiting upon the transition from arrest into growth. Other functions (i.e. translational and mitochondrial) are also found at low levels during recovery. However, as protein abundance depends on transcript levels, transcription is one mechanism to mitigate the overall cellular shortage of growth-promoting proteins. Since operons are enriched for ribosomal and mitochondrial genes, this initial transcription upon transition to growth functions in large part to rapidly generate the translational machinery as well as mitochondrial functions. In addition, as translation bears higher energetic cost than transcription, it is likely that other evolutionary solutions alleviate the limited translational and mitochondrial functions.

In this study we demonstrated how operons can be advantageous to facilitate rapid recovery from arrested states into a growing state when transcriptional resources are limited. However, additional selective factors may also lead to operon formation. For example, the vast majority of operon genes are expressed in the germline (Reinke and Cutter, 2009; Reinke et al., 2004), presumably to support fast proliferation. Expression of growth genes that support proliferation could benefit from operon structures, because less transcriptional resources are required. One option is that the saved resources can then be directed towards proliferation processes and to speed up germline production. In addition, clustering growth genes into operons is presumably not the only mechanism that accelerates their expression during recovery. For example, Pol II pausing on growth genes during arrest may also contribute to rapid transcription during recovery (Baugh et al., 2009), and different animals may have evolved additional, yet undefined strategies. Nevertheless, operons appear to have evolved in diverse metazoans that share rapid recovery from growth arrest as part of their lifecycle.

Experimental Procedures

Estimation of transcriptional speed up due to operons

1. Estimation of the Hill coefficient—While the Hill coefficient for the assembly of transcriptional machinery on promoters is not known, we can estimate its value based on the available data. Transcription rate is often described by the Hill function: $\beta=\beta_{max}\cdot TM^n / (K^n+TM^n)$, or in it's dimensionless form: $\beta/\beta_{max}=(TM/K)^n/(1^n+(TM/K)^n)$, where β is transcription rate, TM/K is the normalized concentration of the transcriptional machinery (K denotes the concentration of TM which enables 50% of maximal transcription rate), and n is the Hill coefficient.

Transcription during dauer state was measured and found to be ~15% of its' maximal value (Dalley and Golomb, 1992). We also find that during the dauer state many components that

make up the transcriptional machinery are found at 50%-60% of their levels as found in growing worms with maximal transcription rate (Fig 2). While the value of K is not known, it is usually estimated to be on the order of the concentration of the TM. Thus, for $K \sim TM$, during dauer state TM/K=50%-60%. Plugging these numbers into Hill function to extract the Hill coefficient, n, results in n= ~3. Analysis of parameter space where TM/K ranges from 40% to 70%, and β/β_{max} ranges between 10%-20% results in n ranging from ~2 to ~6 with a mean value of 3.1.

2. Estimation of the increase in transcription rate—We find that in cases of recovery from both L1 arrest and dauer, ~5,000 genes are up-regulated (by at least 20% during the first 4 hours of recovery). ~1,200 of these significantly up-regulated genes are operon genes. To calculate the possible increase in transcriptional machinery levels due to organization of genes into operons, we note that if there had not been operons, then on average 5,000 RNA Pol II units would be required (assuming 1unit per promoter). However, having ~1,200 genes clustered in ~400 operons (the average size of an operon is ~3) reduces the number of putative RNA Pol II binding promoters by 800, thus effectively increasing the concentration of the transcriptional machinery by ~20%. Since there are many additional operon genes moderately transcribed during the recovery period, we estimate the effective increase in transcriptional machinery levels to be much higher than 20%.

Transcription of polycistronic mRNAs is longer than transcription of mono-cistronic mRNAs. With ~3 genes per operon the average time to transcribe an operon gene is about three times longer than a non-operon gene. However, given that transcription initiation occurs in bursts (Chubb et al., 2006; Golding et al., 2005), the rate limiting step is the assembly of multiple components on promoter sites to initiate transcription (Carey, 1998). Since genes on an operon use the same already-formed transcription complex, the transcription of a gene downstream on an operon can start with no delay. Elongation rate, on the other hand, is presumably not affected since once transcription initiated the rate of elongation is similar whether or not transcriptional machinery is limited. Operons therefore, not only reduce promoter regions that compete for valuable resources, but also reduce the time required for re-initiating transcription. With over 1,000 operons and an average of ~3 genes per operon, the transcription of ~2,000 highly-expressed growth genes takes place with no extra initiation delays. This by itself further increases the transcription rate of operon genes.

In summary, with a Hill coefficient of ~3, TM/K~50%, and at least 20% increase in transcriptional resources, transcription rate increases by at least 70% (see Fig 4A). When considering n=4 (also a plausible value) transcription rate is doubled (100% increase).

Recovery from L1 arrest and growth rate measurements

L1 arrested larvae were prepared as previously described (Baugh et al., 2009). Recovery was initiated after six days of growth arrest in S-basal. Detailed description of the experiments, the mutants, specific media and concentrations used are found in the SI. Worms' growth rate was determined by measuring the worms' length over time. For this, we spread hundreds of worms on a pre-dried NG plate (9 cm) and the plates were image-scanned using an automated stage. Images were analyzed using custom-made Matlab scripts. To measure growth rates upon recovery, we averaged the growth rate during the first ~30 hours of recovery, and for non-arrested worms we averaged the first ~30 hrs after the bleach. During these first ~30 hours, we sampled the worms between 4 to 6 times so that growth rates are averaged over ~5 time points.

Analysis of operon genes in B. *malayi* and *P. pacificus*

The list of *B. malayi* operons was based on (Ghedin et al., 2007). These putative operon genes were inferred to be operonic if transcribed in the same direction and ≤1000 bp apart. We therefore considered various intergenic distance thresholds for assigning genes into operons. We extracted full coordinates from the gff3 file, WS187, WormBase. To determine operon genes in *P. pacificus* we calculated intergenic distances by extracting gene models using a gff2 file downloaded from WormBase (WS197). Detailed bioinformatics procedures are found in SI.

Analysis of gene expression and operon genes in *C. intestinalis*

Gene expression profiles during the lifecycle of *C. intestinalis* is based on Azumi et al (Azumi et al., 2007) and was downloaded from Gene Expression Omnibus (GEO) - http://www.ncbi.nlm.nih.gov/geo. The list of *C. intestinalis* operon genes is based on (Satou et al., 2008) and was downloaded from http://hoya.zool.kyoto-u.ac.jp/download_kh.html. We used GO annotations from blast2go (Conesa et al., 2005) (http://blast2go.bioinfo.cipf.es/annot_euk_others) to retrieve GO annotations of *C. intestinalis*. Detailed analysis is found in SI.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Tom Blumenthal, Erich Schwarz, Hillel Schwartz, Shalev Itzkovitz, Ronen Zaidel-Bar, and Ron Milo for critical review and helpful comments. We thank Elodie Ghedin for sharing the list of operon genes in *B. malayi*, the Mitani laboratory and the Caenorhabditis Genetics Center for deletion strains, WormBase, and the Genome BC *C. elegans* Gene Expression Consortium http://elegans.bcgsc.bc.ca/ for SAGE data (produced at the Michael Smith Genome Sciences Centre with funding from Genome Canada). AZ was supported by the European Molecular Biology Organization (EMBO), the Human Frontier Science Program (HFSP), and the Caltech Center for Biological Circuit Design. LRB was an ACS postdoctoral fellow. PWS is an investigator with the HHMI, which supported this work.

References

Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin EG, Deleury E, Perfus-Barbeoch L, Anthouard V, Artiguenave F, Blok VC, et al. Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. Nat Biotechnol. 2008; 26:909–915. [PubMed: 18660804]

Alon, U. An Introduction to Systems Biology: Design Principles of Biological Circuits. CRC press; 2006. 2007

Azumi K, Sabau SV, Fujie M, Usami T, Koyanagi R, Kawashima T, Fujiwara S, Ogasawara M, Satake M, Nonaka M, et al. Gene expression profile during the life cycle of the urochordate Ciona intestinalis. Dev Biol. 2007; 308:572–582. [PubMed: 17572404]

Barriere A, Felix MA. High local genetic diversity and low outcrossing rate in Caenorhabditis elegans natural populations. Curr Biol. 2005; 15:1176–1184. [PubMed: 16005289]

Baugh LR, DeModena J, Sternberg PW. RNA Pol II Accumulates at Promoters of Growth Genes During Developmental Arrest in C. elegans. Science. 2009; 324:92–94. [PubMed: 19251593]

Baugh LR, Hill AA, Slonim DK, Brown EL, Hunter CP. Composition and dynamics of the Caenorhabditis elegans early embryonic transcriptome. Development. 2003; 130:889–900. [PubMed: 12538516]

Blackwell TK, Walker AK. Transcription mechanisms. WormBook. 2006:1–16. [PubMed: 18050436]

Blumenthal T, Evans D, Link CD, Guffanti A, Lawson D, Thierry-Mieg J, Thierry-Mieg D, Chiu WL, Duke K, Kiraly M, Kim SK. A global analysis of Caenorhabditis elegans operons. Nature. 2002; 417:851–854. [PubMed: 12075352]

Blumenthal T, Gleason KS. Caenorhabditis elegans operons: form and function. Nat Rev Genet. 2003; 4:112–120. [PubMed: 12560808]

Carey M. The enhanceosome and transcriptional synergy. Cell. 1998; 92:5–8. [PubMed: 9489694]

Cassada RC, Russell RL. The dauerlarva, a post-embryonic developmental variant of the nematode Caenorhabditis elegans. Dev Biol. 1975; 46:326–342. [PubMed: 1183723]

Chen N, Stein LD. Conservation and functional significance of gene topology in the genome of Caenorhabditis elegans. Genome Res. 2006; 16:606–617. [PubMed: 16606698]

Chubb JR, Trcek T, Shenoy SM, Singer RH. Transcriptional pulsing of a developmental gene. Curr Biol. 2006; 16:1018–1025. [PubMed: 16713960]

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005; 21:3674–3676. [PubMed: 16081474]

Dalley BK, Golomb M. Gene expression in the Caenorhabditis elegans dauer larva: developmental regulation of Hsp90 and other genes. Dev Biol. 1992; 151:80–90. [PubMed: 1577199]

Dieterich C, Clifton SW, Schuster LN, Chinwalla A, Delehaunty K, Dinkelacker I, Fulton L, Fulton R, Godfrey J, Minx P, et al. The Pristionchus pacificus genome provides a unique perspective on nematode lifestyle and parasitism. Nat Genet. 2008; 40:1193–1198. [PubMed: 18806794]

Ganot P, Kallesoe T, Reinhardt R, Chourrout D, Thompson EM. Spliced-leader RNA trans splicing in a chordate, Oikopleura dioica, with a compact genome. Mol Cell Biol. 2004; 24:7795–7805. [PubMed: 15314184]

Ghedin E, Wang S, Spiro D, Caler E, Zhao Q, Crabtree J, Allen JE, Delcher AL, Guiliano DB, Miranda-Saavedra D, et al. Draft genome of the filarial nematode parasite Brugia malayi. Science. 2007; 317:1756–1760. [PubMed: 17885136]

Golden JW, Riddle DL. The Caenorhabditis elegans dauer larva: developmental effects of pheromone, food, and temperature. Dev Biol. 1984; 102:368–378. [PubMed: 6706004]

Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. Cell. 2005; 123:1025–1036. [PubMed: 16360033]

Guiliano DB, Blaxter ML. Operon conservation and the evolution of trans-splicing in the phylum Nematoda. PLoS Genet. 2006; 2:e198. [PubMed: 17121468]

Hillier LW, Coulson A, Murray JI, Bao Z, Sulston JE, Waterston RH. Genomics in C. elegans: so many genes, such a little worm. Genome Res. 2005; 15:1651–1660. [PubMed: 16339362]

Huang P, Pleasance ED, Maydan JS, Hunt-Newbury R, O'Neil NJ, Mah A, Baillie DL, Marra MA, Moerman DG, Jones SJ. Identification and analysis of internal promoters in Caenorhabditis elegans operons. Genome Res. 2007; 17:1478–1485. [PubMed: 17712020]

Jacob F, Perrin D, Sanchez C, Monod J. Operon: a group of genes with the expression coordinated by an operator. C R Hebd Seances Acad Sci. 1960; 250:1727–1729.

Johnson TE, Mitchell DH, Kline S, Kemal R, Foy J. Arresting development arrests aging in the nematode Caenorhabditis elegans. Mech Ageing Dev. 1984; 28:23–40. [PubMed: 6542614]

Klass M, Hirsh D. Non-ageing developmental variant of Caenorhabditis elegans. Nature. 1976; 260:523–525. [PubMed: 1264206]

Lawrence JG. Gene organization: selection, selfishness, and serendipity. Annu Rev Microbiol. 2003; 57:419–440. [PubMed: 14527286]

Lercher MJ, Blumenthal T, Hurst LD. Coexpression of neighboring genes in Caenorhabditis elegans is mostly due to operons and duplicate genes. Genome Res. 2003; 13:238–243. [PubMed: 12566401]

Lewis K. Persister cells, dormancy and infectious disease. Nat Rev Microbiol. 2007; 5:48–56. [PubMed: 17143318]

Mayer MG, Sommer RJ. Natural variation in Pristionchus pacificus dauer formation reveals cross-preference rather than self-preference of nematode dauer pheromones. Proc Biol Sci. 2011

Opperman CH, Bird DM, Williamson VM, Rokhsar DS, Burke M, Cohn J, Cromer J, Diener S, Gajan J, Graham S, et al. Sequence and genetic map of Meloidogyne hapla: A compact nematode genome for plant parasitism. Proc Natl Acad Sci U S A. 2008; 105:14802–14807. [PubMed: 18809916]

Price MN, Huang KH, Arkin AP, Alm EJ. Operon formation is driven by co-regulation and not by horizontal gene transfer. Genome Res. 2005; 15:809–819. [PubMed: 15930492]

Qian W, Zhang J. Evolutionary dynamics of nematode operons: easy come, slow go. Genome Res. 2008; 18:412–421. [PubMed: 18218978]

Reinke V, Cutter AD. Germline expression influences operon organization in the Caenorhabditis elegans genome. Genetics. 2009; 181:1219–1228. [PubMed: 19204375]

Reinke V, Gil IS, Ward S, Kazmer K. Genome-wide germline-enriched and sex-biased expression profiles in Caenorhabditis elegans. Development. 2004; 131:311–323. [PubMed: 14668411]

Sanford T, Golomb M, Riddle DL. RNA polymerase II from wild type and alpha-amanitin-resistant strains of Caenorhabditis elegans. J Biol Chem. 1983; 258:12804–12809. [PubMed: 6630208]

Satou Y, Mineta K, Ogasawara M, Sasakura Y, Shoguchi E, Ueno K, Yamada L, Matsumoto J, Wasserscheid J, Dewar K, et al. Improved genome assembly and evidence-based global gene model set for the chordate Ciona intestinalis: new insight into intron and operon populations. Genome Biol. 2008; 9:R152. [PubMed: 18854010]

Schrimpf SP, Weiss M, Reiter L, Ahrens CH, Jovanovic M, Malmstrom J, Brunner E, Mohanty S, Lercher MJ, Hunziker PE, et al. Comparative functional analysis of the Caenorhabditis elegans and Drosophila melanogaster proteomes. PLoS Biol. 2009; 7:e48. [PubMed: 19260763]

Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. Interdependence of cell growth and gene expression: origins and consequences. Science. 2010; 330:1099–1102. [PubMed: 21097934]

Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A, et al. The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS Biol. 2003; 1:E45. [PubMed: 14624247]

Wang J, Kim SK. Global analysis of dauer gene expression in Caenorhabditis elegans. Development. 2003; 130:1621–1634. [PubMed: 12620986]

Whittle CM, McClinic KN, Ercan S, Zhang X, Green RD, Kelly WG, Lieb JD. The genomic distribution and function of histone variant HTZ-1 during C. elegans embryogenesis. PLoS Genet. 2008; 4:e1000187. [PubMed: 18787694]

Zaslaver A, Kaplan S, Bren A, Jinich A, Mayo A, Dekel E, Alon U, Itzkovitz S. Invariant distribution of promoter activities in Escherichia coli. PLoS Comput Biol. 2009; 5:e1000545. [PubMed: 19851443]

Zorio DA, Cheng NN, Blumenthal T, Spieth J. Operons as a common form of chromosomal organization in C. elegans. Nature. 1994; 372:270–272. [PubMed: 7969472]

Highlights

- Recovery from growth arrested states requires up regulation of thousands of genes.
- Transcriptional resources are one limiting factor during the transition from arrest to growth.
- Clustering growth genes into operons can increase transcription and recovery rates.
- This evolutionary solution is found in various animals and can explain why operons evolved in metazoans.

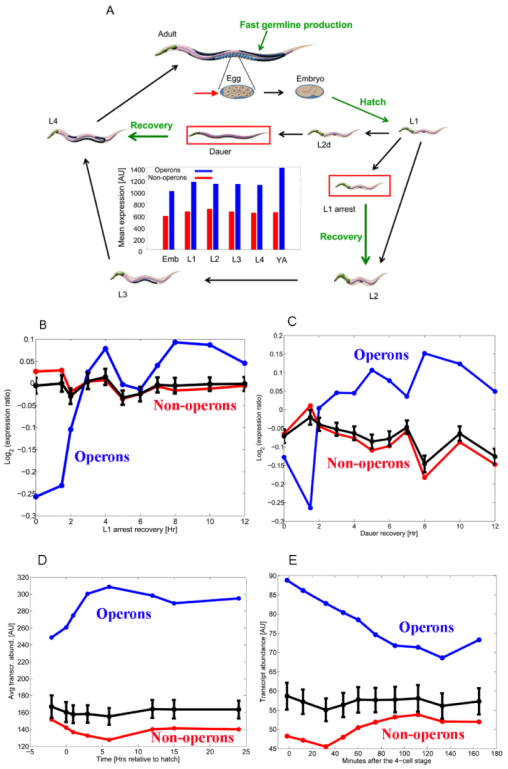

Figure 1. Expression profile of operon genes is anti-correlated to the expression profile of operon genes

(**A**) The lifecycle of free-living nematodes (*e.g., C. elegans*) consist of four larval stages followed by an adult stage. If unfavorable conditions arise during larval development worms stop growing and arrest at the L1 state, or proceed through L2d stage into dauer, a highly-resistant and long-lived state. When conditions improve the worms recover and resume normal development. Fast growing stages are denoted by green arrows. The red arrow marks the first divisions of the egg, a process characterized by fine-regulation rather than rapid growth and mass accumulation. Red rectangles indicate a growth arrested state. (Inset) The average expression levels of operon genes are higher compared to non-operon genes. The higher expression level is consistent throughout all developmental stages. Average expression levels calculated based on microarray data obtained from http://elegans.bcgsc.bc.ca/. Emb, embryo; YA, young adult. (**B-E**) Expression patterns of operon genes (blue) and non-operon genes (red). Bootstrap analysis is in black. a set of random genes (with the same number of operon genes as that included in the data set) was pooled from the total set of genes (including operon genes) and their average calculated. This process was iterated 1,000 times, and the average is plotted with error bars (s.e.m). (**B**) Time-series during recovery from L1 arrest. (**C**) Time-series during recovery from dauer state. (**D**) Time-series following hatching. (**E**) Time-series during early embryogenesis. See also Table S1 and Fig S1-S3.

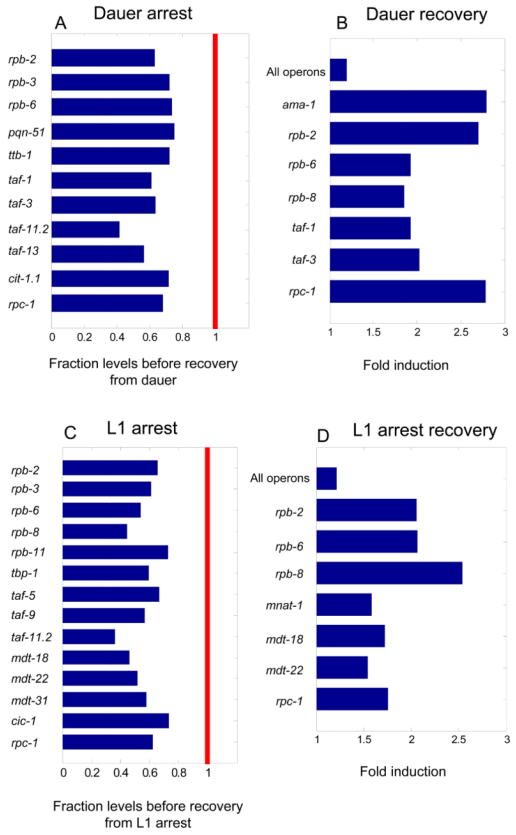

Figure 2. Growth arrested worms have low levels of transcriptional resources and are therefore sensitive to any further reduction

(**A,C**) Many genes associated with transcriptional machinery are found at lower levels (~50%) during growth arrest when compared to their levels in non-arrested worms (Wang and Kim, 2003); in particular (**A**) during dauer arrest and (**C**) during L1 arrest. Shown are the genes that are at most 75% of maximum levels. Many other transcriptional machinery genes are found at moderately lower levels. (**B,D**) Expression of many transcriptional machinery genes is significantly increased upon recovery. (**B**) Fold change in expression levels following dauer recovery. (**D**) Fold change in expression levels following recovery from L1 arrest. Fold change is calculated as the ratio between the mean expressions during 3-5 hrs of recovery to the mean expression during the first 1.5 hrs of recovery. Note that while the average increase in expression levels for all operon genes is ~20% expression levels of many transcriptional genes is increased by 100% and more. The list of transcriptional machinery genes is based on (Blackwell and Walker, 2006).

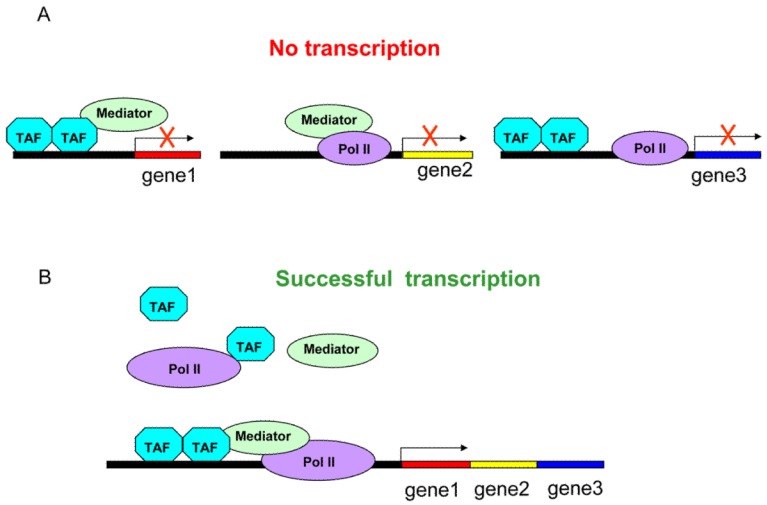

Figure 3. Clustering genes into operons is beneficial when transcriptional resources are limiting

An illustration demonstrating how operons become advantageous for successful transcription when transcriptional resources are limited. Assume there is a fixed amount of transcriptional resources (for example, 2 RNA Pol II, 2 mediator complexes and 4 TAF complexes) in the cell and that a functional transcriptional initiation complex requires half of these resources (1 RNA Pol II, 1 mediator and 2 TAFs). Suppose also there are three growth genes required to be expressed in high levels. **(A)** The three genes are on separate monocistronic units each having its own promoter. In this case the probability to form a functional transcriptional initiation complex is low as transcriptional resources are stochastically distributed among the different binding sites of the different promoters. **(B)** The three genes are on the same operon regulated by a single shared promoter. In this case, the number of potential biding sites competing for transcriptional resources is reduced by three fold. This effectively increases the concentration of the transcriptional resources increasing the probability to form productive transcriptional initiation complex.

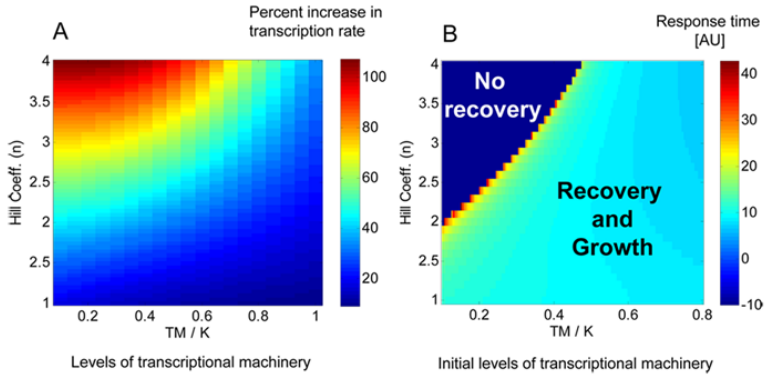

Figure 4. A model based on Hill–function analysis demonstrating that a moderate increase or decrease in transcriptional resources can lead to recovery or death

(A) Heat map showing the fold increase in transcription rate obtained if levels of transcriptional machinery are increased by 20%. Considerable enhancement in transcription rate is found for a wide range of possible values of transcriptional machinery levels (*TM/K*) and Hill-coefficient values (*n*). **(B)** Heat map demonstrating that response time and recovery are very sensitive to initial levels of transcription machinery. The upper-left blue area indicates that recovery is impossible. A sharp, high boundary is found for intermediate levels of transcriptional machinery. Thus, any small increase in the levels of the transcriptional machinery can push the animals from the no recovery regime to the recovery and growth regime. Note that the maximal possible increase in transcription rate (A) is obtained at the 'poor' no recovery area shown in blue in (B). See also Fig S6.

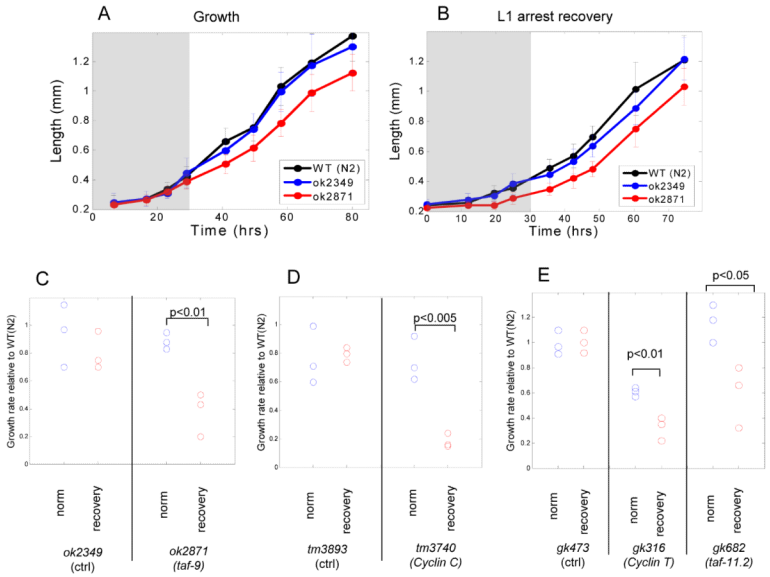

Figure 5. Recovery of arrested worms is significantly impaired by reduced levels of transcriptional components

Deletion mutants lacking transcriptional machinery components were analyzed for their growth rate during normal growth and during recovery from L1 arrest. Growth rates were calculated based on the first ~30 hours following recovery, or during the first ~30 hours of non-arrested growing L1 larvae (indicated by the gray-shadowed area). **(A-B)** An example of growth curves (measuring length of the worms) over ~80 hrs following: **(A)** hatching into non-arresting conditions, or **(B)** following recovery from L1 arrest. The three strains given as an example are: *taf-9(ok2871)*, wild type (N2), and a control strain *gpa-16(ok2349)*, defective in a non transcriptional-related gene. **(C-E)** Three experimental repeats measuring the relative growth rates of the different deletion mutants: **(C)** *taf-9(ok2871)* and *gpa-16(ok2349)*. **(D)** Cyclin C (*tm3740*) and the control *srbc-58(tm3893)*, a serpentine receptor. **(E)** Cyclin T (*gk316*), *taf-11.2(gk682)*, and the control *miR-84(gk473)*. Relative growth rates were calculated by taking the ratio between the growth rates of the mutants and the growth rate of N2 (wild type), which was included in each of the growth experiments. Error bars indicate standard error of the mean. P-values are based on t-tests. Each time point is the mean of ~100 worms.

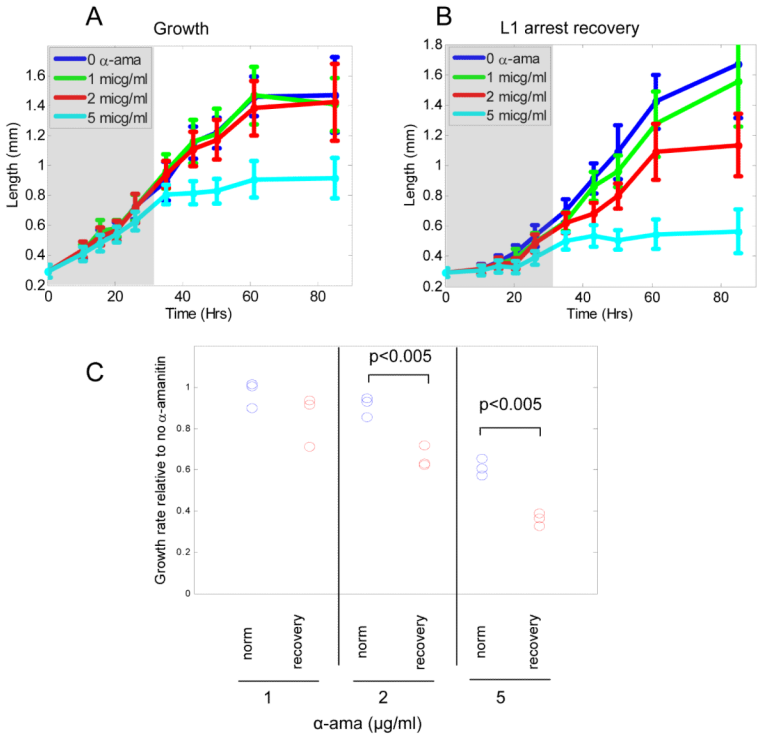

Figure 6. Recovery of arrested worms is significantly impaired by reduced levels of RNA Pol II

Growth rate of wild-type (N2) worms was measured in the presence of RNA Pol II inhibitor α-amanitin. **(A)** Growth curves in non-arresting conditions from L1 stage to adults. (**B**) Growth curves of arrested L1 worms following recovery. At time zero arrested L1 larvae were washed and allowed to recover in the presence of the indicated levels of α-amanitin. At the same time actively growing L1 larvae were treated with the same concentrations of α–amanitin. **(C)** Summary of three repeats measuring relative growth rates in the presence of α-amanitin. Growth rates were calculated by averaging growth during the first ~30 hrs following the addition of α-amanitin (indicated by the gray-shadowed area). Relative growth rates are the ratio between growth rates in the presence of α-amanitin and the absence of α-amanitin. Error bars indicate standard error of the mean. P-values are based on t-tests. Each time point is the mean of ~100 worms.

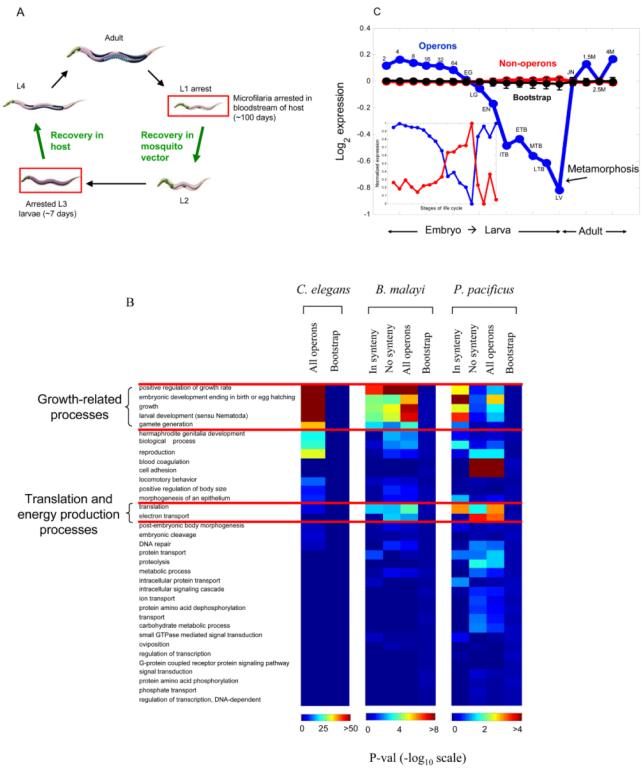

Figure 7. Various metazoans contain operons that are enriched with growth genes and may promote rapid recovery form developmental arrested states

(**A**) Lifecycle of the parasitic nematode *B. malayi*. *B. malayi* passages between mosquitoes, which serve as vectors, and humans, the infected hosts. Microfilaria arrest as L1s in the human bloodstream. They are then taken by feeding mosquitoes where they recover and develop until they arrest in the L3 stage. The next time mosquitoes feed on human blood the L3 arrested worms are transmitted to the bloodstream where they recover and resume growth until they reach adult stage and produce progeny (microfilaria). Red rectangles denote growth arrested states. (**B**) Operon genes in *B. malayi* and *P. pacificus* typically comprise growth related genes. P-values are calculated by hyper geometric test (HGT) performed on GO annotations (WS190). Included are GO annotations that contain more than 200 genes. Bootstrap analysis verified that the set of genes in GO annotations is not biased to contain preferentially growth related genes. Data are clustered using hierarchical clustering algorithm (euclidean distances, average linkage). (**C**) Microarray time-series during the lifecycle of *Ciona intestinalis*. Average expression of operon genes is in blue, and average expression of non-operon genes is in red. For bootstrap analysis (black) a set of random genes (with the same number of operon genes included in the data set) was pooled from the total set of genes (including operon genes) and their average was calculated. This process was iterated 1,000 times and the average is plotted with error bars (s.e.m). Since expression of operon genes ranges much widely than expression of non-operon genes, the two expression profiles were normalized (inset). Notations used: 2-cell, 4-cell…64-cell embryos; EG – early gastrula; LG – late gastrula; EN – early neurula; ITB – initial tailbud; MTB middle tailbud; LTB – late tailbud; LV – early larvae; JN – juvenile; 1.5M – 1.5 months-old adults etc. See also Fig S7.